*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

**Big Data y Machine Learning para Economía Aplicada**

**Taller 1**

**2023-02**

**Link del repositorio en Github:**

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

Angie Ariza Quitian - 201325848

Andrés Felipe Diaz Barreto - 200610686

**Presentado por**

Duvan Javier Farfán López – 201317299

Daniel Eduardo Sandoval - 200712968

• Introducción. La introducción expone brevemente el problema y si existen antecedentes. Describe brevemente los datos y su idoneidad para abordar la pregunta del conjunto de problemas. Contiene un avance de los resultados y las conclusiones principales.

• Datos. En este conjunto de problemas, se requiere que amplíes las variables en tus datos (recuerda ampliar los datos de entrenamiento y prueba). Como mínimo, debes agregar seis variables adicionales:

– Al menos 4 predictores procedentes de fuentes externas; estos pueden ser de mapas de Open Street.

– Al menos 2 predictores procedentes del título o la descripción de las propiedades.

Cuando escribas esta sección, debes:

Describir los datos, su idoneidad para el problema y el proceso de construcción de la muestra, incluyendo cómo se limpiaron los datos, se combinaron y se crearon nuevas variables.

Incluir un análisis descriptivo de los datos. Como mínimo, debes incluir una tabla de estadísticas descriptivas y dos mapas con su interpretación. Sin embargo, se espera un análisis profundo que ayude al lector a comprender los datos, su variación y la justificación de tus elecciones de datos. Utiliza tu conocimiento profesional para agregar valor a esta sección. No lo presentes como una "lista seca" de ingredientes.

• Modelo y Resultados. Esta sección presenta el modelo con la mejor puntuación enviado para su evaluación. Cuando escribas esta sección, incluye:

– Una explicación de las variables utilizadas para entrenar este modelo, recuerda utilizar las variables que agregaste en la sección anterior.

– Una explicación detallada de cómo se entrenó, la selección de hiperparámetros y cualquier otra información relevante.

De acuerdo con la literatura académica relacional con los precios de las viviendas, este se suele determinar a través del modelo de precios hedónicos, que busca estimar el valor de un bien inmueble a través de las características que poseen bienes que diferenciados (Taylor, L. 200X). Es así que, surgen variables o características de los inmuebles como la cercanía equipamientos (hospitales, colegios, parques, estaciones de policía, etc.), así como el área de la propiedad y otra serie de características (ubicación).

Para este ejercicio se determinaron como variables internas a las características de las viviendas, las siguientes variables: tipo de la vivienda (casa o apartamento), el número de habitaciones (dormitorios), si cuenta o no con parqueadero/garaje, el área (tamaño) de la propiedad, el número de baños y si cuenta o no con deposito.

**Tipo de vivienda**

Se refiere a si es una casa o un apartamento. Bogotá es una de las ciudades con mayor densidad en América Latina, por lo que el mercado inmobiliario ha venido cambiando en el que la construcción de apartamentos predomina desde hace varios años.

**Habitaciones**

– Una sección que compare el rendimiento de la presentación con la mejor puntuación con otras presentaciones enviadas a Kaggle. Estas presentaciones deben incluir especificaciones entrenadas utilizando modelos de Regresión Lineal, Ridge, Lasso, Elastic Net, CART, Random Forest y Boosting. Indica en tu archivo de presentación el nombre del modelo utilizado.

• Conclusiones y recomendaciones. En esta sección, expones brevemente las conclusiones principales de tu trabajo.